心理科学进展 2022, Vol. 30, No. 8, 1715–1733
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2022.01715

# 新世纪 20 年国内结构方程模型方法研究与模型发展*

王 阳[1] 温忠麟[2] 李 伟[3,4] 方 杰[5]

([1]广东金融学院公共管理学院，广州 510521) ([2]华南师范大学心理学院/心理应用研究中心，广州 510631)
([3]西北民族大学教育科学与技术学院，兰州 730124) ([4]华中师范大学人工智能教育学部，武汉 430079)
([5]广东财经大学新发展研究院/应用心理学系，广州 510320)

**摘 要** 新世纪前 20 年，国内结构方程模型(SEM)方法研究主要涉及 5 个主题：模型发展、参数估计、模型评价、测量不变性及特殊数据处理，特别是模型发展方面(即 SEM 的各种变式)有较多成果。对每个主题，在简述背景知识的基础上，系统总结了方法学研究发展及成果。最后也讨论了 SEM 的国外方法学研究进展和未来研究方向。

**关键词** 结构方程模型，模型发展，参数估计，模型评价，测量不变性
**分类号** B841

## 1 引言

结构方程模型(structural equation model, SEM)是回归模型的推广，有回归模型不具备的诸多优势：可以同时处理多个自变量和因变量，满足社科研究中理论模型日益复杂化的需求；可以同时分析显变量和潜变量，符合社科研究中变量普遍具有内隐性的特点；允许自变量有测量误差，参数估计精度更高；具有丰富的拟合评价指标来评价模型等。这些优势使 SEM 成为社科研究中重要的统计方法。国内 SEM 方法研究最早见于张建平(1993)对 SEM 的评介文章。侯杰泰等(2004)出版国内第一本 SEM 专著大大推动了 SEM 在国内的传播和应用。

新世纪以来国内 SEM 方法学研究取得长足进展，成果甚丰。以中国知网(https://www.cnki.net/)全文数据库为数据源，出版年限设为 2001~2020 年，关键词包括：结构方程、潜变量、隐变量、结构模型、测量模型、验证性因子分析、验证性因素分析、线性结构关系、协方差结构、协方差矩阵，经筛查得到期刊发表的 SEM 论文 192 篇，所属学科和发表年份见表 1。不计 SEM 入门知识及学科应用现状的文章，也不计以应用为主旨的文章。

从学科来看，多达 13 个学科发表过 SEM 方法研究，其中心理学发文最多，其次是医学和统计学。从出版时间来看，2001~2005 年 18 篇，2006~2010 年 50 篇，2011~2015 年 68 篇，2016~2020 年 56 篇。SEM 方法学研究总体上呈上升趋势，2011~2015 年是高峰期。这与国内心理统计方法研究发展趋势一致(温忠麟 等, 2021)。

根据内容所涉及的 SEM 研究主题对文章进行分类。有 10 篇以上文章讨论同一主题的归为一类，否则归为“其它”类。这样，国内(期刊)的 SEM 方法学研究论文可以分为 5 个主题：模型发展(69 篇)、参数估计(38 篇)、模型评价(17 篇)、测量不变性(15 篇)、特殊数据处理(10 篇)，还有其它 43 篇。这与温忠麟等(2021)对 SEM 研究主题的分类略有出入，因为本文不限于心理学期刊发表的论文。本文总结新世纪 20 年(2001~2020 年)国内(期刊) SEM 方法研究进展，并通过对比近年来国外 SEM 方法研究，了解该领域一些前沿论题。

收稿日期：2021-12-28
* 国家自然科学基金项目(32171091)、广东省哲学社会科学规划项目(青年项目) (GD21YXL04)、国家社会科学基金项目(17BTJ035, 19BMZ080)、甘肃省教育科学规划项目(GS[2021]GHB1777)、广东省普通高校创新团队项目(人文社科) (2019WCXTD005)、广东省教育科学规划项目(2020GXJK342)和中央高校基本科研业务费专项资金项目(31920210120)资助。
通信作者：温忠麟, E-mail: wenzl@scnu.edu.cn

表 1 2001~2020 年国内(期刊)各学科发表 SEM 方法论文的频数分布(按发文篇数排序)

| 刊物类别 | 年份 | | | | 合计 |
|---|---|---|---|---|---|
| | 2001~2005 | 2006~2010 | 2011~2015 | 2016~2020 | |
| 心理学 | 5 | 12 | 28 | 25 | 70 |
| 医学 | 2 | 12 | 14 | 12 | 40 |
| 统计学 | 5 | 13 | 12 | 3 | 33 |
| 其它学科 | 2 | 8 | 11 | 10 | 31 |
| 综合性刊物 | 4 | 5 | 3 | 6 | 18 |
| 合计 | 18 | 50 | 68 | 56 | 192 |

注：其它学科包括系统科学(7 篇)、管理学(6 篇)、数学(4 篇)、体育学(3 篇)、信息学(3 篇)、教育学(3 篇)、社会学(2 篇)、生物学(1 篇)、计算机科学(1 篇)和语言学(1 篇)。综合性刊物主要是高校学报。表中涉及基于 SEM 的信度计算和中介调节方法论文 29 篇，由于这些文章另有专文综述(见：方杰 等, 2022; 温忠麟, 陈虹熹 等, 2022; 温忠麟, 方杰 等, 2022)，这里只做篇数统计，正文未出现。

## 2 结构方程模型的发展

结构方程模型包含测量模型(measurement model)和结构模型(structural model)。测量模型反映潜变量及其测量指标的关系，单独使用即为验证性因子分析(confirmatory factor analysis, CFA)；结构模型反映(潜)变量间的影响关系，如果结构模型中的潜变量换成测量指标的均分或总分进行分析，就是路径分析(path analysis)。近年来, SEM 发展出不少新变式，测量模型方面主要有双因子模型、探索性结构方程模型、特殊设计的测量模型(随机截距因子分析模型、预设路径模型及瑟斯顿模型)和形成性测量模型；而结构模型方面主要是主客体互依模型；全模型方面(即同时包含测量和结构模型的完整 SEM)主要是测量指标合并(即条目打包)的 SEM。此外, SEM 在群体异质性研究和追踪研究中的发展也值得关注。

### 2.1 测量模型的发展

#### 2.1.1 双因子模型

传统 CFA 中，一个问卷条目只在一个因子上有非零载荷。双因子模型(bi-factor model)则在传统多因子 CFA 模型基础上，允许问卷条目额外负载于一个全局因子(顾红磊 等, 2014)，全局因子可以是特质因子(可用于探索和验证高阶因子结构、计算基于 CFA 模型的信度及分析各类因子与效标变量的关系)或方法因子(可用于检验共同方法偏差)。

双因子模型和高阶因子模型有嵌套关系，任何一个高阶因子模型都可以转换为一个双因子模型，但只有满足比例约束(即每个维度中全局因子载荷和局部因子载荷之比为常数)的双因子模型才能转换为一个高阶因子模型(顾红磊 等, 2014)。模拟研究发现，对于满足比例约束的特殊双因子模型，其对潜效标变量的预测精度不如高阶因子模型(徐霜雪 等, 2017)；而在不满足比例约束的更一般情况下，使用双因子模型进行预测效度分析时，模型拟合、统计检验力和效度系数估计精度优于高阶因子模型(温忠麟 等, 2019)。

#### 2.1.2 探索性结构方程模型

传统 CFA 模型是典型的独立分群模型，问卷条目在非目标因子上载荷固定为 0，这可能造成因子间相关虚高，也容易使 CFA 拟合不了探索性因子分析得到的因子结构。探索性结构方程模型(exploratory structural equation model, ESEM)可以有效弥补这些局限。ESEM 在 CFA 的基础上，允许跨因子载荷不为 0。更真实地体现因子结构的同时，还容易获得较好的拟合。麦玉娇和温忠麟(2013)详细介绍了 ESEM 的原理，比较了 ESEM 和探索性因子分析、CFA 的异同点，并给出了使用 ESEM 的建议。值得一提的是，如果模型参数估计采用贝叶斯法并设定特殊先验，则不仅能像 ESEM 一样放宽跨因子载荷限制，还能更灵活地放宽残差相关限制(Muthén & Asparouhov, 2012)。

#### 2.1.3 特殊设计的测量模型

**随机截距因子分析模型。**随机截距因子分析模型(random intercept factor analysis model)在一般 CFA 模型的基础上，增加一个潜截距因子。潜截距因子取值在被试间有变异，在条目间无变异

(即潜截距因子在所有条目上载荷固定为常数), 以此反映被试的某种稳定特质(如社会赞许性或默许偏差), 可以用于解释和控制条目表述效应。研究发现, 相比于双因子模型, 随机截距因子分析模型有助于提升特质变异在问卷分数总变异中的占比, 使问卷在存在条目表述效应的条件下具备更好的结构效度, 特质变异大于方法变异(韦嘉 等, 2016)。

**预设路径模型**。预设路径模型(fixed-links model)专门用于实验研究, 主要作用是分离实验中的目标因子和无关因子, 令实验目标概念的测量更准确。潜变量由 1 个目标因子和若干个非目标因子构成。因子载荷固定, 且允许跨因子载荷。目标因子载荷根据已有理论知识和经验指定, 而非目标因子载荷统一设定为一个常数。模型的评价除了拟合指数, 还依赖目标因子方差。目标因子方差显著表示该潜变量所代表的心理过程是完成任务所必须的(任学柱 等, 2017)。

**瑟斯顿模型**。社科研究中常见到配对比较任务和排序任务。配对比较任务如: 在配对呈现的两个面孔图片中选择更喜欢的一个; 而排序任务如: 将 3 个面孔图片按喜好程度排序。任务产生的数据为等级数据, 不满足经典测量理论的基本假设(王珊 等, 2014), 最好用专门的模型分析, 比如瑟斯顿模型(Thurstone model)。该模型中潜变量表示比较或排序任务中的某个选项, 而潜变量所属测量指标为被试对该选项与其它选项的偏好选择结果。若被试偏好当前选项, 则因子载荷固定为 1, 否则为−1。对于排序任务, 测量指标的残差设定为 0。这种模型的优势在于, 可以同时得到各任务选项均值差异的详细信息及被试的个体差异信息; 能侦测到细小的选项差异; 可以避免社会赞许性的影响(宋晓娟, 刘红云, 2016)。

#### 2.1.4　形成性测量模型

形成性测量模型(formative measurement model, FM)是测量模型的一种特殊形式。FM 与传统的测量模型(也可称为反映性模型; reflective measurement model, RM)的主要区别体现在(贾跃千, 宝贡敏, 2009): (1) RM 中因子影响测量指标; FM 则相反, 因子由测量指标所构建; (2) RM 要求测量指标具有高内部一致性, 可以互换; FM 测量指标可以不相关甚至负相关, 编制测量指标更要紧的是涵盖构念的所有方面, RM 的信效度及拟合评价指标往往不适用于 FM; (3) RM 的误差项体现在测量指标层次; FM 的误差项体现在潜变量层次; (4) RM 适用于验证性的研究, 注重理论模型与实际数据的拟合及参数估计的准确性; 而 FM 适用于探索性的研究, 更关注测量指标对潜变量的预测效力。

虽然 FM 并未如 RM 一样流行(贾跃千, 宝贡敏, 2009), 但在国内也受到一定关注, 有一些研究介绍该模型的基本原理和特征(王念新 等, 2013; 王晓丽 等, 2011; 叶浩生, 李明, 2014)。王念新等(2011)的模拟研究表明, 如果将 FM 误设为 RM 可能造成路径系数的估计偏差和两类错误率的升高, 建议利用模型细化法和模型分解法来避免模型误设。

### 2.2　结构模型的发展

社科研究常常关注成对数据(dyadic data), 如配偶双方、师生双方及上下级双方等在同一变量上的数据。此类数据往往不具备独立性(李育辉, 黄飞, 2010)。为避免可能的两类错误膨胀, 需要专门的统计方法。主客体互依模型(actor-partner interdependence model, APIM)是一种专门分析成对变量关系的结构模型。以教师和学生的共情能力(自变量)对教师和学生的领悟社会支持(因变量)的影响为例, 主客体互依模型具体路径设定包括 4 个部分: (1)学生/教师共情对其自身领悟社会支持的影响, 即主体效应; (2)学生/教师共情对对方的领悟社会支持的影响, 即客体效应; (3)学生和教师共情的相关, 这一设定可以在分析某一自变量的效应时, 控制另一自变量的影响; (4)学生和教师的因变量残差相关, 用于控制除自变量之外的其它因变量互依性来源(刘畅, 伍新春, 2017)。通过 APIM 可以分析主体效应和客体效应的大小和方向特点及何者更占优势。APIM 具体的原理(李育辉, 黄飞, 2010)、分析流程(刘畅, 伍新春, 2017)、模型变式(如含中介和调节变量的 APIM; 陈莎 等, 2020; 刘畅, 伍新春, 2017)及软件操作(如 Mplus 和 SPSS; 陈莎 等, 2020; 何娟 等, 2018)可参考相关方法文献。

### 2.3　全模型的发展

条目打包(item parceling)可以视为全模型构建的一种特殊方法。这种方法将测验原始条目聚合成条目包, 牺牲了测量模型分析的可靠性, 但能改善结构模型的参数估计和模型拟合。国内关

于条目打包的研究涉及两个方面，一是对打包技术及策略的介绍和演示。如卞冉等(2007)详细介绍了条目打包的基本逻辑、优劣势及具体方法。而吴艳和温忠麟(2011)在此基础上进一步细化了打包方法并给出了操作流程。对于单维量表，前一组研究者从经济实用的角度出发，推荐随机法打包(即条目打包不依循任何规律；卞冉 等，2007)；后一组研究者从拟合提升最大的角度出发，推荐平衡法(即首先将条目按因子载荷高低排序，再按照“S”形顺序分配各包条目的方法；吴艳，温忠麟，2011)；对于多维量表，两组研究者都建议采用内部一致性法(即每个维度内所有条目聚合成一个条目包；卞冉 等, 2007；吴艳，温忠麟, 2011)，因为该法保留了潜变量的多维结构。王若宾等(2014)建议对全模型中的多维量表按维度均分打包，并通过实例演示说明该方法可以简化模型，提升对路径系数的检验力，并获得理想拟合。第二个方面是条目打包和其它模型形式的比较，如杨彤骥等(2010)用一个应用实例比较了条目打包、路径分析及未打包的全模型，结果发现条目打包的模型拟合指数比较好，而未打包全模型所得的测定系数 $R^2$ 要高于条目打包和路径分析。

### 2.4 群体异质性研究中的结构方程模型

探究不可直接观测的群体异质性是很多研究的兴趣所在，由此催生了大量基于 SEM 的异质性分析方法。刘源和刘红云(2015)及李丽霞等(2015)对此类方法进行了总结，主要包括：潜类别/剖面模型、因子混合模型和多水平潜类别模型。

#### 2.4.1 潜类别/潜剖面模型

潜类别/潜剖面模型是根据外显测量指标对被试进行分类的测量模型(张洁婷 等，2010)，是聚类分析思想在 SEM 中的体现。若测量指标是类别变量，则为潜类别模型(latent class model, LCM)；若是连续变量，则为潜剖面模型(latent profile model, LPM)。

国内关于 LCM/LPM 的研究涉及三个方面。第一是对 LCM/LPM 基本原理和分析流程的介绍。如张洁婷等(2010)介绍了 LCM 的统计原理、分析过程及在心理学研究中的应用情况。郭小玲等(2009)、孟灿等(2010)及曾宪华等(2013)分别用模拟研究数据和应用案例演示了 LCM 的分析流程。尹奎等(2020)介绍了 LPM 的基本原理、步骤及在组织行为研究中的应用。

第二是对包含协变量的 LCM/LPM 或潜类别与协变量关系分析的研究，如王孟成和毕向阳(2018)总结了含协变量的 LCM (即回归混合模型)的分析方法并给出了 Mplus 语句模板。学界一般推荐用 LTB 法(Lanza et al., 2013)处理分类结果变量，用 BCH 法(Bolck et al., 2004)或稳健三步法处理连续结果变量，后者也适用于协变量是预测变量的情形(王孟成，毕向阳, 2018)。张洁婷等(2017)介绍了 LPM 后续分析(即分出潜类别后，探究潜类别与前因后果变量关系的分析)的方法，并通过模拟研究比较指出，纳入式分类分析法(即潜剖面分类时就将后续分析需要的变量作为协变量纳入模型)的后续分析参数估计效果比较好，兼有准确性和稳健性。进一步研究则发现，当纳入的后续分析变量包含结果变量及自变量与潜类别变量的乘积项时，后续参数估计效果比较好。更具体的分析流程见张洁婷等(2019)。

第三是通过模拟研究考察 LCM/LPM 的分类效果，包括 LCM/LPM 与其它个体中心方法的比较及不同数据和模型条件对分类结果的影响。如马文超等(2014)通过模拟比较发现，除只有两个潜类别且各类样本容量极端不均衡的特殊情况外，LCM 和快速聚类法及混合 Rasch 模型的分类精度相当。赵丽等(2013)通过模拟比较指出 LPM 的分类准确率高于系统聚类法。王孟成、邓俏文、毕向阳等(2017)通过模拟研究考察了 LPM 中类别数、类间距、样本容量和测量指标个数对 Entropy 分类精确性的影响。

#### 2.4.2 因子混合模型

因子混合模型(factor mixture model)将 CFA 和 LCM 整合在同一个模型中，可以看作是一种以 CFA 中潜变量为分类指标的 LCM；也可以看作是考虑了群体异质性的 CFA 模型。陈宇帅等(2015)介绍了因子混合模型的基本原理、主要优势、应用方向及分析步骤。李观海等(2020)比较了不同样本量和因子间相关条件下潜类别因子模型(因子混合模型变式之一，假定群体异质性仅存在于潜均值)和 LCM 的表现，结果显示前者的模型拟合及分类效果都好于后者，且模型更精简和易于识别。

#### 2.4.3 多水平潜类别模型

多水平 LCM 是专门针对多水平结构数据的 LCM，可以针对相同的测量指标在个体(层 1)和

组织(层 2)层级分别进行分类。张洁婷等(2013)介绍了多水平潜类别模型的基本原理，以小学英语能力测验为例演示了多水平 LCM 的分析步骤，并比较了多水平 LCM 与一般 LCM 的效果差异。

### 2.5　追踪研究中的结构方程模型

追踪研究是一种通过对相同的研究对象和变量重复测量多次，对序列数据进行处理、分析，以了解变量发展趋势、变量之间相互关系及个体差异的研究设计。国内追踪研究中的 SEM 方法主要涉及描述发展趋势及差异的模型(潜增长模型、多阶段增长模型、潜类别增长模型、增长混合模型、多阶段增长混合模型和潜在转变模型)及探究变量间相互影响的模型(交叉滞后模型)。

#### 2.5.1　描述发展趋势及差异的模型

**潜增长模型。**潜增长模型(latent growth model, LGM)以不同时间点的变量观测值作为测量指标，用一个截距因子反映被试的特质基线水平(因子载荷固定为 1)，若干斜率因子反映特质的线性或非线性变化趋势。可同时用于探究心理特质的个体差异与发展趋势。已有不少方法文献介绍 LGM 的基本原理、常见变式、软件操作(如 Mplus 和 SAS)及优缺点等(李丽霞 等, 2012; 宋秋月，伍亚舟，2017; 苏荣海，徐茂洲，2017; 许碧云 等，2007)。李丽霞等(2014)从模型数学形式、前提假设、数据格式、参数估计及建模的灵活性与复杂性等角度比较了 LGM 和多水平模型，指出多水平建模更加简单直接；而 LGM 更加灵活，可以放松对测量误差相等的限制，自由估计每次测量误差，参数估计精度更好。

**多阶段增长模型。**LGM 假定个体的发展轨迹总是连续的，忽视发展可能存在的阶段性与转折点(如前期增长慢而后期增长快)。多阶段增长模型(piecewise growth models, PGM)允许增长曲线有不同发展阶段。刘源等(2013)通过模拟研究探讨了 SEM 和多水平模型框架下定义的 PGM 的参数估计效果差异及将 PGM 错误设定为无阶段模型的后果。

**潜类别增长模型和增长混合模型。**LGM 假设潜变量在个体间的发展轨迹相同，忽视了可能存在的异质性。潜类别增长模型(latent class growth model, LCGM)和增长混合模型(growth mixture model, GMM)通过结合 LCM 和 LGM，可以将个体按照特质发展趋势的不同进行分类，前者假定同一类别内个体无差异，后者无此限制(李丽霞等, 2015)，前者可视为后者特例(王孟成 等, 2014; 肖健 等, 2020)。有关 LCGM 的详细介绍参见吕浥尘和赵然(2018)或王孟成等(2014)；有关 GMM 的详细介绍参见刘红云(2007)、王孟成等(2014)、肖健等(2020)或喻嘉宏等(2018)。

**多阶段增长混合模型。**多阶段增长混合模型(piecewise growth mixture model, PGMM)是 PGM 和 GMM 的结合，允许发展轨迹既分阶段，又有群体异质性。王婧等(2017)详细描述了 PGMM 的基本原理、常见模型形式、参数估计方法及影响因素、样本量需求、拟合评价指标、应用现状及未来研究方向。刘源等(2014)则利用模拟研究考察了潜类别间距和发展模型形态对 PGMM 的模型选择及参数估计的影响。

**潜在转变模型。**潜在转变模型(latent transition model, LTM)是 LCM 在追踪研究中的推广，不仅可以探究各时间点可能存在的潜类别，允许潜类别发生变化，还可以考察个体从某一类别转变为其它类别的概率。王碧瑶等(2015)和黄明明(2019)介绍了 LTM 的统计原理，并分别以青少年冲动行为和英语阅读理解测验为例，演示了 LTM 的分析过程和结果解读。黄明明(2018)介绍了基于混合项目反应理论的 LTM 的理论基础、转变机制、模型特性、应用现状及发展前景。

#### 2.5.2　探究变量间相互影响的模型

交叉滞后模型(cross-lagged model)通过分析多变量间跨时间的相互影响，可以探索变量间谁是因谁是果。模型重点关注的效应包括(刘源 等, 2022)：(1)自回归效应，即同一变量前测对后测的影响，反映变量的跨时间稳定性(重测信度)；(2)交叉滞后效应，即控制变量 *B* 前测时变量 *A* 前测对变量 *B* 后测的效应和控制变量 *A* 前测时变量 *B* 前测对变量 *A* 后测的效应。*A* 和 *B* 的因果顺序判断依据是前测原因变量对后测结果变量的预测作用(以标准化路径系数表示)应明显高于前测结果变量对后测原因变量的预测作用(周广帅 等, 2020)。该方法可以使问卷研究更好地满足因果推论对于因果先后顺序及无关变量控制的要求(温忠麟, 2017)。

## 3　结构方程模型参数估计方法

SEM 的参数估计基于对协方差结构的分析。

设 **Σ**(**θ**)和 ***S*** 分别为根据理论模型得出的协方差矩阵和样本协方差矩阵, **θ** 为参数向量。用拟合函数 *F*[***S***, **Σ**(**θ**)]表示 **Σ**(**θ**)与 ***S*** 的距离。参数估计的过程即求得使 *F*[***S***, **Σ**(**θ**)]达到最小值的 **θ** 估计值的过程。不同拟合函数构建方式产生了不同的参数估计方法。使用最多的方法是极大似然估计(maximum likelihood estimate, ML)。当数据非正态, 通常使用稳健极大似然估计(robust maximum likelihood estimate, MLR)或均值和方差校正的加权最小二乘估计(weighted least squares with mean and variance adjusted, WLSMV)。前者对因子间相关及参数标准误的估计更准确, 后者对因子载荷的估计更准确(Li, 2016), 且更适合计分点数偏少的数据。

国内有关 SEM 参数估计方法的研究, 主要涉及两个方面：一是方法的介绍, 主要包括偏最小二乘法(partial least square, PLS)和贝叶斯法(Bayesian methods); 二是参数估计方法的比较。

### 3.1 偏最小二乘法

传统参数估计方法求解 SEM 是令拟合函数最小化的过程, 强调参数估计精度; 而 PLS 求解 SEM 则是令残差方差最小化的过程, 强调方程中预测变量对结果变量的预测精度, 这一特点与形成性建模的主要目的(追求测量指标对因子的解释能力最大化)非常契合(王念新 等, 2013), 因此 PLS 常常被用于分析形成性模型, 通常使用专用软件如 SmartPLS、semPLS 或 WarpPLS 进行分析。相比于传统参数估计方法, PLS 的优势主要在于：(1)更适合小样本和非正态数据; (2)适合复杂模型(即变量数与样本容量之比较高的模型); (3)适合形成性模型分析; (4)适合对多个预测变量的作用进行探索的 SEM 研究(骆雷, 2020)。

国内有关 PLS 的介绍主要涉及两个方面, 一是对 PLS 的介绍与评价, 如朱利平和刘莉(2005)、宁禄乔等(2007)详细介绍了 PLS 的参数估计过程。孙继红和杨晓江(2009)总结并讨论了基于 PLS 的 SEM 涉及的三个重要问题：测量模型如何选择、如何用 bootstrap 法估计和检验参数及如何评价模型。刘金兰等(2005)讨论了 PLS 算法的几何意义。宁禄乔和刘金兰(2007)通过模拟研究考察了 PLS 估计 SEM 参数的效果, 发现该方法低估结构系数, 高估因子载荷。当样本容量较大时, 结果基本可信。

第二是 PLS 在特定模型及数据中的拓展, 如林盛等(2006)介绍了两个潜变量的 PLS 算法如何扩展到适用于多个潜变量。程豪和易丹辉(2016)介绍了用 PLS 估计二阶因子模型的基本原理和优点; 赵萍(2011)讨论了使用 PLS 估计二阶因子模型时不同权重估计算法模式的优劣。王芝皓等(2020)介绍了 PLS 路径模型的分位效应(一般的 SEM 是用自变量预测因变量均值, 分位效应则是指用自变量预测因变量百分位数)估计方法。孟洁和王惠文(2009)、李顺勇和岳利梅(2017)分别介绍了如何将基于 PLS 的路径分析和基于 PLS 的 SEM 用于处理成分数据(即在 0~1 之间取值、和为 1 的数据)。任红梅和王绥(2010)介绍并推荐使用模糊 PLS 来为具有不确定性的数据(比如 Likert 量表的中间项往往表示不确定)进行 SEM 建模。应用实例表明, 该方法的参数估计精确度优于未经模糊处理的 PLS, 且模型拟合更好。

### 3.2 贝叶斯法

贝叶斯统计分析是一种将已有的对于待估计参数的经验和知识(即先验信息)融入参数估计的过程。贝叶斯法用于 SEM 时的参数估计过程如下：

(1)首先设定理论模型, 这一步和普通 SEM 无异。(2)设定所有未知参数的先验分布参数(也称超参数)。研究中感兴趣的参数一般是目标因子载荷和路径系数, 通常假定服从正态分布, 需设定均值和方差两个参数, 可参照已有研究特别是元分析(晏宁, 毛志雄 等, 2018)。 需要强调, 尽管很多软件能给出默认先验分布参数(无信息先验, 相当于仅采用贝叶斯的估计框架, 但却不利用先验信息), 但贝叶斯法的核心优势是利用先验信息帮助参数估计。有研究发现使用无信息先验并不比 ML 强, 甚至小样本时基于无信息先验的贝叶斯估计还可能造成严重不稳定和有偏的估计结果(Smid & Winter, 2020)。因此采用贝叶斯法时, 应尽量使用有信息先验。

(3)利用马尔可夫链蒙特卡洛方法(Markov chain Monte Carlo, MCMC)的 Gibbs 抽样获取参数后验分布(张言彩, 2009)。估计结果是否收敛可以利用潜在尺度缩减因子(potential scale reduction)、踪迹图(trace plots)及有效样本容量(effective sample size)等判定(具体的解读方法参见：王孟成, 邓俏文, 毕向阳, 2017 和 Smid & Winter, 2020)。估计收敛后, 后验分布的集中量数(如均值或中位数)可以作为模型参数的点估计值。同时, 有了后

验分布, 可以获取基于贝叶斯法的可信区间(credible interval)。

相较传统的基于频率理论的 SEM 参数估计方法, 贝叶斯法的主要优点是: (1)小样本条件下估计效果更好, 尤其是能够提供有效先验信息时; (2)计算速度更快(王孟成, 邓俏文, 毕向阳, 2017); (3)当模型复杂, 或待估计参数和样本容量之比较高时, 传统参数估计容易出现收敛问题, 而贝叶斯方法往往能收敛到恰当解(梁莘娅, 杨艳云, 2016; 王孟成, 邓俏文, 毕向阳, 2017); (4)贝叶斯方法受非正态的影响更小(晏宁, 李英 等, 2018); (5)传统方法不可识别的模型(比如令所有跨因子载荷和残差相关自由估计会因用尽自由度导致模型不可识别), 通过纳入先验信息和采用 MCMC 方法, 仍有可能识别(晏宁, 李英 等, 2018); (6)通过贝叶斯法获得的可信区间解释上比传统置信区间更直观。

王孟成、邓俏文和毕向阳(2017)、晏宁和李英等(2018)介绍了基于贝叶斯法的 SEM (BSEM)的基本概念并用示例演示了分析过程和结果解释。张沥今等(2019)分析了贝叶斯法在不同 SEM 变式(如普通的测量模型、潜中介模型、潜增长模型、多组 SEM 及多层 SEM)中的应用优势及现状, 并对 BSEM 的模型评价及可用软件做了介绍。秦正积等(2020)介绍了如何使用 SAS 软件调用 OpenBUGS 程序以更高效地实现 BSEM。

### 3.3　其它参数估计方法

除了 PLS 和贝叶斯法, 吴瑞林(2012)建议用 Tikhonov 正则化方法修正 ML 的参数估计, 模拟研究表明这有助于提升收敛率和收敛速度, 减少不恰当解, 并降低估计偏差。对于二分类数据和顺序等级数据, 有研究者建议采用多项相关矩阵来估计模型参数(吴瑞林, 祖霁云, 2010), 并给出 LISREL 实例语句(王欢 等, 2012; 吴宇驹 等, 2012)。比起基于皮尔逊相关矩阵的参数估计, 多项相关矩阵得到的参数估计值偏差更小(周映雪 等, 2013)。还有研究介绍 SEM 的约束最小二乘解, 并将其推广到高阶测量模型和多组 SEM 中(童乔凌, 刘天桢 等, 2009; 童乔凌, 邹雪城 等, 2009)。该方法可以提升参数估计的收敛率和收敛速度, 并获得唯一解。

### 3.4　参数估计方法比较

SEM 参数估计方法比较研究一部分关注加权最小二乘法(weighted least squares)、对角加权最小二乘法(diagonal weighted least squares)和广义最小二乘法(generalized least squares)的表现(焦辛妮, 王长义 等, 2015; 焦辛妮, 汪东伟 等, 2015; 吴瑞林, 2010)。不过这些方法应用不多, 且早有文献指出, 与 ML 相比它们的综合表现并不突出(侯杰泰 等, 2004)。

更值得注意的是当前流行的估计方法和新方法的比较。比如刘红云、骆方等(2012)模拟比较了 MLR、WLSMV 和 MCMC 方法估计二分类数据测量模型的精度。结果发现, 三种方法对因子载荷等参数的估计精度都比较好, MLR 和 WLSMV 略优于 MCMC 方法, WLSMV 还具备运算速度快的优势。田晓明和傅珏生(2004, 2005)模拟比较了 ML 和贝叶斯法在 SEM 参数估计中的表现, 结果发现贝叶斯法估计精度略优于 ML, 但并不明显。梁莘娅和杨艳云(2016)模拟比较了 MLR 和无先验贝叶斯法在测量模型中的表现。结果发现: (1)对于识别错误模型的能力, 在非正态条件下, 贝叶斯方法更强; 但在正态条件下, 贝叶斯方法不如 MLR。(2)贝叶斯法收敛能力远强于 MLR, 这个优势在复杂模型(如双因子模型)中尤为明显。晏宁和毛志雄等(2018)通过一个应用实例比较了 ML 和贝叶斯法在小样本潜变量建模中的表现, 尽管二者对路径系数及因子载荷等参数估计结果接近, 但前者会出方差为负这样的异常解。

霍映宝(2006)、李晓鸿(2012)和张军(2007)从参数估计目的、基本原理、前提条件等理论方面比较了 PLS 与传统参数估计方法(如 ML、加权最小二乘法、对角加权最小二乘法和广义最小二乘法)的区别; 章刚勇(2015)则对二者做了模拟比较, 结果发现 PLS 参数估计的稳定性强于传统方法, 但是对误设模型的敏感性较低。刘燕和陈英武(2007)提出用广义最大熵法(Generalized Maximum Entropy)估计 SEM 参数, 并模拟比较了广义最大熵法和 PLS 的参数估计偏差, 结果广义最大熵法获得的参数均方误差总是小于 PLS。

## 4　结构方程模型评价

对 SEM 参数的解读是建立在假设模型与实际数据拟合良好的基础上的。评价模型拟合主要通过拟合指数, 涉及到的方法学问题包括新拟合指数的提出、拟合指数的临界值、拟合指数的选

择及拟合指数之外的模型评价标准。此外，多个模型间的比较和选择也是模型评价的重要方面。

### 4.1 拟合指数的发展

自 SEM 流行以来，研究者陆续提出过 40 多个拟合指数，温忠麟等(2004)探讨了好的拟合指数应有的性质(不受样本容量系统影响、惩罚复杂模型及对误设模型敏感)。现在普遍认为，CFI (comparative fit index)、TLI (Tucker-Lewis index)、RMSEA (root mean square error of approximation)和 SRMR (standardized root mean square residual)是统计学性质较好的拟合指数(温忠麟，刘红云，2020)。

近年来有研究尝试提出新的拟合指数，比如王凯等(2018)针对拟合指数 GFI (goodness-of-fit index)受样本容量系统影响及未惩罚复杂模型的缺陷，提出了校正后的 GFI (corrected GFI, CGFI)。另一种新拟合指数是 Yuan 和 Chan (2016)提出的基于等效性检验的拟合指数(equivalence-testing-based fit indice; 王阳 等, 2020)。其基本思想是，针对传统 $\chi^2$ 的逻辑问题(将零假设模型完全拟合作为想要证明的假设)，设置新的零假设(模型误设大于一个可以容忍的小正数)和备择假设(模型误设不大于可容忍误设)。在此基础上，提出与之相适应的拟合指数 $RMSEA_t$ 和 $CFI_t$。与传统 RMSEA 和 CFI 不同，$RMSEA_t$ 和 $CFI_t$ 具有推断统计性质。以 $RMSEA_t$ 为例，它表示模型误设的大小不超过 $RMSEA_t$，且做出这一推断犯错误的概率不超过显著性水平 α。

另外，对于近年来受到关注的 BSEM，有专门的拟合评价指标：(1)后验预测 $p$ 值(posterior predictive $p$ value, $ppp$)。它反映观测数据拟合函数和基于后验分布的样本数据拟合函数间的差距(梁莘娅，杨艳云，2016)，近似 0.5 表示模型拟合良好，接近 0 或 1 表示拟合不佳。(2)贝叶斯因子(Bayes factors)。可粗略地理解为当前数据对两个竞争模型支持强度的比值。经验上，两个竞争模型的贝叶斯因子大于 10，表示有较强证据支持贝叶斯因子分子所代表的模型；贝叶斯因子小于 1/10，表示有较强证据支持贝叶斯因子分母所代表的模型(胡传鹏 等, 2018)；贝叶斯因子在 1/3~3 之间，表示数据对两个模型的支持力度差不多(张沥今 等, 2019)。(3) $L_\nu$。该统计量相较于贝叶斯因子，具备计算量更小，且对先验信息依赖更少的优点。$L_\nu$ 越小的模型拟合越好。李云仙等的系列论文详细介绍了 $L_\nu$ 的统计原理，并将其用于带有有序变量和缺失数据的 SEM 及两水平 SEM 的模型选择(李云仙，王学仁，2011, 2012；李云仙，杨爱军，2014)。(4)贝叶斯信息指数(Bayesian information criterion, BIC)和异常信息指数(deviance information criterion, DIC)。这两个指标均只用于模型比较，数值越小表示模型拟合越好。

此外，当数据不服从正态分布，有研究提出用 Satorra-Bentler 校正 $\chi^2$ 评价模型拟合(刘小琴 等, 2013)。金勇进和梁燕(2005)介绍了适用于基于 PLS 的 SEM 的拟合评价指标，包括因子共同度、$R^2$ (外生潜变量对内生潜变量的预测效果)、冗余(由外生潜变量所解释的内生潜变量测量指标平均方差)等。

### 4.2 拟合指数的临界值

针对常用拟合指数 CFI、TLI、RMSEA 和 SRMR，一般认为，CFI 和 TLI 不低于 0.9 (Bentler & Bonett, 1980)、RMSEA 和 SRMR 不高于 0.08 时(Browne & Cudeck, 1992)，模型可以接受。但也有研究者建议使用更高的标准：CFI 和 TLI 不低于 0.95、RMSEA 和 SRMR 不高于 0.05 (Hu & Bentler, 1999)，现在通常将这一较严苛的标准作为模型拟合优秀的标准。此类临界标准只是一种基于经验的约定俗成的判断。

有研究者通过设计一些特殊的真模型和错误模型展开模拟研究，根据各拟合指数采取不同临界值时两类错误率之和来确定不同条件下最佳的拟合临界值(郭庆科 等, 2007, 2008)。然而，真模型与错误模型的差距复杂多样，拟合最佳临界值会随这一差距的变化而改变。有文献指出这种通过模拟研究确定固定临界值的做法是不恰当的(温忠麟，侯杰泰, 2008; Marsh et al., 2004)。

### 4.3 拟合指数的选用

由于存在诸多拟合指数可供使用，需要考虑选择报告哪些指数。王长义等(2010)通过总结和分析已有的拟合指数性能评估研究，认为 TLI 和 RMSEA 是最值得信赖的拟合指数，不过 CFI、RNI、Mc、SRMR 和 $\chi^2/df$ 也有一定参考价值。还有研究者通过分析拟合指数的公式，发现除了理论模型不优于独立模型(即只有显变量且互不相关的模型)的极端情况，总有 CFI≥TLI (温涵，梁韵斯, 2015)。所以，当 TLI 可以接受，CFI 就是多

余的。而且, CFI 不惩罚复杂模型(温忠麟 等, 2004), 当几个模型都能拟合数据时, 无法帮助研究者选出更简洁的模型。这样, 研究者选择报告 TLI、RMSEA 和 SRMR 即可。

### 4.4　其它模型评价标准

应用研究者往往将拟合指数作为最重要甚至唯一的模型评价标准。很多文献对此做法提出批评。一方面, 每种指数都只是从一个特定角度评价拟合, 有其固有局限(王阳 等, 2020)。另一方面, 拟合指数普遍受模型拟合程度之外的其它因素影响, 如样本容量、数据分布、因子载荷、参数估计方法等(王阳 等, 2020; 温忠麟, 侯杰泰, 2008; Shi & Maydeu-Olivares, 2020)。所以评价模型拟合有必要参考其它标准, 如参数估计评价、整体拟合、内部拟合及复核效度检验等(侯杰泰 等, 2004; 温忠麟 等, 2004, 温忠麟, 侯杰泰, 2008; 郑文智, 吴文毅, 2014)。

首先, 可考察参数估计过程是否正常收敛。出现识别或收敛问题往往是因为模型设定不合理。然后检查 4 个方面内容：(1)模型参数符号是否恰当、是否统计显著？不显著考虑修改(侯杰泰 等, 2004)。(2) $R^2$ 是否足够大？对于测量模型, 太小的 $R^2$ 说明载荷太低, 意味着条目信度过低(温忠麟, 侯杰泰, 2008)。不少研究发现拟合指数倾向于支持信度更低的模型(如 Greiff & Heene, 2017), 所以需要同时权衡拟合指数和信度。(3)残差矩阵有无异常元素(温忠麟, 侯杰泰, 2008), 太大的残差绝对值意味着理论模型和数据有明显差距。(4)哪些路径或载荷有较大的修正指数(侯杰泰 等, 2004)？过大的修正指数意味着模型中的箭头可能指错了位置。看拟合指数之前先分析这些与参数估计有关的评价指标可以提升对误设模型的检验力(温忠麟, 侯杰泰, 2008; 郑文智, 吴文毅, 2014)。

参数估计评价之后, 才看拟合指数。拟合指数是对模型整体拟合情况的评价。除此之外, 还可以考察模型内部拟合, 即从模型的内在质量评价每个潜变量的设置是否恰当(郑文智, 吴文毅, 2014)。主要包括(1)检查测量工具信度, 可以使用 CFA 计算合成信度; (2)检查测量工具效度, 包括内容效度、效标关联效度和结构效度。

此外, 还可以考察复合效度(即交叉验证), 将数据一分为二, 用样本一(校正样本)估计参数, 然后将这些参数赋给样本二, 并查看其拟合。还可以对比两个样本所得结果的差异(郑文智, 吴文毅, 2014)。理想情况是两个样本所得拟合都比较好且接近。

### 4.5　模型比较和选择的策略

对于同一批数据, 能够良好拟合的模型可能不止一个, 此时需要通过比较多个模型的拟合来选择最优模型。柳恒超等(2007)介绍了嵌套模型的概念和特点及嵌套模型和非嵌套模型的比较和选择方法。对于嵌套模型, 他们建议比较 5 个模型：独立零模型 Mn、饱和模型 Ma、感兴趣的理论模型 Mt, 以及两个其次感兴趣的理论模型 Mc 和 Mu。首先用 Ma 的 $\chi^2$ (最小 $\chi^2$)和 Mn 的自由度(最大自由度)进行 $\chi^2$ 检验, 如果统计显著, 则所有模型都不可接受; 如果不显著, 则通过各模型间的 $\chi^2$ 差异检验寻找最优模型。对于非嵌套模型, 一般建议比较期望交叉验证指数(expected cross-validation index, ECVI)和 Akaike 信息指数(Akaike information criterion, AIC), 这两个指数越小的模型越好。骆方和张厚粲(2006)以创造性思维测验为例, 演示了如何根据 $\chi^2$ 差异检验和模型的简洁性来比较和选择最佳 CFA 模型。此外, 郭芸(2005)介绍了如何在非线性 SEM 中利用贝叶斯因子选择最佳模型。

## 5　多组测量模型(测量不变性)

测量不变性(measurement invariance)是指 SEM 在不同群体或时间有相同结构及参数值, 一般涉及到检验模型形态、因子载荷、截距项、因子与误差的方差−协方差及潜均值不变性(潜均值比较涉及均值结构模型, 是 SEM 的特殊形式之一; 侯杰泰 等, 2004)。国内大部分有关测量不变性的方法文献侧重于介绍测量不变性要检验的各种模型、检验流程及模型评价标准(例如：白新文, 陈毅文, 2004; 刘军, 2005; 刘军, 吴维库, 2005; 刘砚燕, 袁长蓉, 2015; 魏修建, 郑广文, 2015; 武淑琴, 张岩波, 2011; 武淑琴 等, 2009; 许宏晨, 2010; 张连生 等, 2012; 赵必华, 2007; 郑广文 等, 2014)。此外, 还有两个测量不变性的研究方向受到关注。第一是特定模型或数据中如何实现测量不变性分析。如郑显亮等(2011)介绍了二阶因子模型的测量不变性分析方法。其与传统测量不变性分析的主要差别在于, 各项不变性约束都要分

别在一阶和二阶水平上进行设定。李冲和刘红云(2011)建议在进行等级数据的测量不变性分析时，使用 WLSMV 估计参数，并以基于 WLSMV 的校正 $\chi^2$ 差值检验法(correct chi-square difference test, DIFFTEST)比较嵌套模型。他们的模拟研究表明，WLSMV 对因子载荷和阈值参数的估计准确性高，DIFFTEST 的两类错误率也可以接受(李冲，刘红云，2011)，且这些方法的表现并不逊于基于项目反应理论的方法(刘红云，李冲 等, 2012)。

第二方面，测量不变性分析中，不少约束模型太过严格，不易实现(温聪聪 等, 2018)。有研究专门讨论这一问题。比如，当形态不变性和载荷不变性成立，但截距不变性未能满足，如果研究者还想比较潜均值，可以采用投影法(王阳 等, 2020; Deng & Yuan, 2016)。该法将每个组的显变量均值分解为两个正交的成分：公分数(代表潜均值)和特殊因子。这两个成分的跨组不变性分析都不依赖截距项，从而使潜均值比较可以直接绕过截距不变这一传统前提。

如果只有形态不变性成立，但之后的不变性约束模型拟合不佳，建议采用对齐法，构建一个反映各组截距和载荷差值的损失函数，并求取使损失函数最小的载荷与截距估计值。此时，约束模型中各组对应参数虽然不完全相等，但足够接近；而且，该模型与形态不变模型有相同的拟合度。对齐法可以视为一种近似不变性模型。该法的具体原理、应用实例和 Mplus 语句参见温聪聪等(2018)。此外，最近也有研究者建议利用 BSEM 分析测量不变性。通过给需要检验的参数设置均值为 0 且方差极小的先验分布，达到放宽参数跨组不变限制，实现近似不变性分析的目的(宋琼雅 等, 2021)。

## 6　结构方程模型中的特殊数据处理

这里说的特殊数据处理主要包括缺失数据、非连续数据、非正态数据及潜变量得分等。

### 6.1　SEM 中的数据缺失问题

SEM 建模时如果缺失数据过多，可能导致模型估计出现问题(如协方差矩阵非正定；林小鹏 等, 2010)。推荐好的缺失数据填补方法及对比不同方法的效果是重要的研究工作。方法推荐方面，多重填补法(multiple imputation, MI)和全信息极大似然估计法(full information maximum likelihood, FIML)是目前最为研究者所推崇的缺失处理方法(王孟成，邓俏文, 2016)。MI 对缺失数据进行多次填补，用每个填补后的完整数据集进行分析从而获得目标参数的多个估计值，最后将多个估计值进行汇总获得最终参数估计值; FIML 则不对缺失值进行替换，而是根据未缺失数据的信息采用迭代的方式进行估计。MI 优势包括充分考虑了数据的不确定性、能更灵活地处理同时包含连续和非连续变量的混合数据(Mansolf et al., 2020)；不足在于 MI 的分析过程更复杂，耗时更长(叶素静 等, 2014)。FIML 的主要优势是操作简便(王孟成，邓俏文，2016)，不足是有时会遇到第一类错误率偏高及模型收敛问题(Mansolf et al., 2020)。

方法比较方面，杨林山和曹亦薇(2012)模拟比较了完全贝叶斯法(将缺失数据视为未知参数，通过模拟全体变量与缺失值的联合后验分布来估计模型参数和缺失值)和部分贝叶斯法(相当于贝叶斯估计时忽略缺失)处理 LGM 中缺失数据的效果。结果发现缺失比例超过 50%时，前者的均方误差明显小于后者，即参数估计精度更高；而缺失比例较小时，二者效果接近。陈楠和刘红云(2015)模拟比较了 ML 和 Diggle-Kenward 选择模型这两种方法处理 LGM 中非随机缺失的效果，总体上后者效果更好。邓居敏等(2018)用一个实证数据比较了列删法(listwise deletion)、期望最大化算法(expectation maximization algorithm)、MI 和 FIML 处理 SEM 中缺失数据的效果，发现 MI 和 FIML 获得的模型拟合较好，期望最大化算法获得的参数标准误最小，而列删法获得的参数标准误最大。王孟成和邓俏文(2016)通过模拟研究探讨了利用 FIML 处理 SEM 数据缺失时辅助变量的作用，总的来说纳入辅助变量有助于得到更可靠的参数估计。

### 6.2　SEM 中的非连续及非正态数据问题

社科研究中所用的问卷数据经常是有序分类的，直接将其作为连续变量建模可能会降低参数估计精度和模型拟合，高静(2012)提出了一个有序分类数据的连续化处理程序，并通过应用示例展示了这种方法对模型拟合的提升。

方敏和黄正峰(2010)介绍了非正态数据的常用 SEM 建模方法，推荐使用 Bollen-Stine Bootstrap 法校正 $\chi^2$ 检验结果，并给出了 Amos 操作演示。

### 6.3 SEM 中的潜变量得分问题

张岩波等(2005)介绍了 SEM 中潜变量得分的估计原理, 并用 LISREL 演示了潜变量得分的计算。刘玥和刘红云(2017)介绍了双因子模型中全局因子和局部因子的加权因子分计算方法, 模拟研究表明相对于其它测验总分和维度分的合成方法, 加权因子分最接近真分数, 信度最高。此外, 张洁婷等(2012)介绍了潜变量数据类型的确定方法。

## 7 其它议题

除上述 5 大主题, 国内 SEM 方法研究还有一些值得关注的主题, 包括传统统计和测量方法如何纳入 SEM 框架、对误差相关问题的讨论、SEM 建模步骤的改进、SEM 检验力分析等。

鉴于 SEM 的诸多优势及其与其它统计方法的兼容性, 有研究者介绍了传统统计和测量方法与 SEM 的结合方法, 以提升分析准确性和建模灵活性。如基于 CFA 模型的信度计算(见: 温忠麟, 陈虹熹 等, 2022)、基于 SEM 的各类中介模型(见: 温忠麟, 方杰 等, 2022)与调节模型(见: 方杰 等, 2022)、基于 SEM 的多水平模型(毕向阳, 2019; 方杰 等, 2011; 张岩波 等, 2006; 张岩波 等, 2008)、基于 SEM 的元分析(桂裕亮 等, 2016; 钱刘兰 等, 2015)、基于 SEM 的时间序列分析(朱苗苗, 2016)、基于 SEM 的人工神经网络模型(颜波 等, 2019; 赵海峰, 万迪昉, 2003)、基于 SEM 的条目因子分析(item factor analysis, 吴瑞林, 涂冬波, 2013)、基于 SEM 的指标体系(即为抽象概念建立的带权重、分层级的测量指标系统; 贾新明, 2011; 斯介生 等, 2014; 田飞, 2007; 王惠文, 付凌晖, 2004; 俞立平, 2020; 张瑛, 王惠文, 2008)、基于 SEM 的模糊综合评价算法(庄伟卿, 刘震宇, 2013)、基于 SEM 的隐马尔可夫模型(王坤 等, 2018; 夏业茂 等, 2016)及基于双因子模型的计算机自适应测验(刘馨婷 等, 2019; 毛秀珍 等, 2019; 毛秀珍 等, 2018)。

通常 SEM 中如无充分理据, 不应设定测量误差相关。当误差项之间存在系统性相关来源时, 可以设定误差相关, 如相同方法测量的条目的误差、同一条目的多次重复测量的误差设定为相关是合理的。另外, 当模型存在多个较大的误差相关修正指数, 可以通过增设潜在共同方法因子来解释误差相关(胡鹏 等, 2018)。

针对 SEM 建模过程中不恰当的模型设定可能降低模型拟合的问题, 陈明亮(2004)提出改进 SEM 的建模步骤, 主要是增加识别和剔除不恰当问卷条目(与所测变量相关偏低或与其它条目相关过大的条目)及相关性过强的变量, 并重新评估测量模型和结构模型拟合的步骤。

安敏(2016)总结了 SEM 中检验力分析的常用方法, 并以师生关系问卷为例说明了如何用 Mplus 进行检验力分析。王惠文和张瑛(2007)介绍了 SEM 的后推预测算法, 即如何根据变量当前形成的 SEM 预测未来变量间的关系, 并通过模拟研究支持了该算法的预测效果。单娜和张笑笑(2020)通过数学推导, 给出了因变量和协变量(同时影响自变量和因变量的无关变量)均为潜变量, 而自变量是显变量时, 平均因果效应的计算方法。贾新明和刘亮(2008)从变量、样本、数据、参数估计和模型解释几个方面比较了 SEM 与联立方程模型的异同点。

## 8 讨论与拓展

新世纪 20 年来, SEM 得到越来越多社会学科的重视和应用, 这也带动了 SEM 方法研究的发展。在国内, 十余个不同学科为 SEM 方法研究做出了贡献, 在 5 个主题上有较多的研究成果。2020 年后至今, 仅追踪研究中的 SEM 这一个方向, 就增加了多篇文献(方俊燕 等, 印刷中; 高文阳 等, 2021; 刘源, 2021; 温聪聪, 朱红, 2021; 袁帅 等, 2021; 郑舒方 等, 2021)。具体介绍见刘源等(2022), 这里不再赘述。在其它方向, 如测量不变性、SEM 中的检验力分析和分类数据处理以及不同类型测量指标(即形成性和反应性指标)的选用策略与相应模型的统计分析方法, 也有新的进展(勾建伟, 夏业茂, 2022; 郭晟豪, 萧鸣政, 2022; 宋琼雅 等, 2021; 翟宏堃 等, 2022)。为了更好地对 SEM 的当下研究有所了解, 这里也介绍一些国外较新的 SEM 方法学研究, 可以发现一些值得国内方法学者未来探索和拓展的方向。

### 8.1 原有主题的拓展

在模型的发展方面, 简要介绍双因子 ESEM 和 SEM 树。尽管双因子模型和 ESEM 各自克服了传统 CFA 的一些重要局限, 但也都有不足, 前者忽视了跨因子载荷普遍存在的事实, 可能使因子相关被高估; 后者则忽视了高阶因子存在的可能

性，容易导致跨因子载荷偏高。将双因子模型和ESEM结合起来的双因子ESEM很好地弥补了两种模型独立使用时的局限(Morin et al., 2016)。

SEM树是SEM和决策树的结合，可以根据研究者选定的预测变量(如儿童的运动技能、学习方式和生活常识等)为某一结果变量或结果变量的变化轨迹(如儿童阅读能力的发展轨迹)进行分类。当存在很多有价值的预测变量、预测变量间普遍存在交互作用或对分类的个数缺乏先验知识时，SEM树被认为是有限混合模型(即LCM、LPM、LCGM、GMM及PGMM这一类方法)的良好替代(Jacobucci et al., 2017)。

在参数估计方法方面，简要介绍模型存在误设时参数估计精确度研究的进展及参数估计不确定性的评价。Lai和Zhang (2017)通过模拟比较研究发现，对于CFA模型，ML在模型有较严重误设时仍可以提供可靠参数点估计；但对于完整的SEM，当误设较大时，各类参数估计方法的偏差都比较大。Pek和Wu (2018)提出的可替换参数估计(fungible parameter estimates)评价了SEM参数估计无关抽样变异的敏感性。

在模型评价方面，简要介绍一些拟合指数(包括局部拟合检验和BSEM的拟合指数)及临界值的新动态。传统的拟合指数只用于评价模型整体，Thoemmes等(2018)提出的局部拟合检验为模型的不同部分分别提供拟合评价，有助于定位模型误设的来源，即使模型未识别或不收敛，依然可以使用。

随着BSEM的流行，如何评价它的拟合优度成为一个重要问题。BSEM使用的拟合评价指标并不是研究者所熟悉的传统拟合指数，而且大多只用于模型比较(如贝叶斯因子和异常信息指数)。Garnier-Villarreal和Jorgensen (2020)用参数后验均值取代$\chi^2$构建了基于BSEM的RMSEA、CFI和TLI等7个新拟合指数。

对于临界值的划分，McNeish和Wolf (2021)提出了一种基于数据模拟技术的动态拟合指数(dynamic fit index)临界值，这种临界值考虑到了影响模型拟合的各种模型和数据特征因素，可以有效地拒绝误设模型。

在测量不变性分析方面，通过定义新的条目功能差异(differential item functioning, DIF)或将不变性问题看成是有关模型参数的聚类或调节问题，可以从新的不同角度阐释跨组测量不变性(Bauer, 2017; de Roover et al, 2020; Schulze & Pohl, 2021)。

此外，还有特殊数据的SEM的拓展，涉及Logistic潜增长模型、名义变量的因子分析等非连续数据的SEM建模方法(Asparouhov & Muthén, 2021)、小样本数据建模相关的研究(Jiang & Yuan, 2017; Smid & Winter, 2020)。

### 8.2 开拓新主题

SEM领域新的主题不断涌现，例如，以探索为目的的SEM、基于实验研究的SEM、工具变量(instrumental variables)在SEM中的使用及SEM软件包的开发等。

SEM绝大多数应用都是验证性质的，但是对于包含海量数据的大型研究，可能需要在缺乏预先假设的前提下探索多个前因变量中哪些是有实际作用的。这种探索视角下的SEM可以用正则化方法(Regularization)来实现。正则化SEM通过给拟合函数增加一个惩罚项或给参数设定特殊的先验(如小方差的跨因子载荷先验)，将偏小的系数缩减到0，从而起到变量或路径筛选的作用(Jacobucci et al., 2016; Lu et al., 2016; Muthén & Asparouhov, 2012; Pan et al., 2017)。

近年来国外不少方法文献将实验研究中的传统统计方法与潜变量建模思路相结合，形成了基于实验研究的潜变量建模方法。例如，Breitsohl (2019)比较了被试间方差分析和两种基于SEM的方法：结构均值模型(structured-means-modeling)和多指标多因模型(multiple-indicator multiple-cause)。两种SEM方法都将因变量表示为潜变量，区别是结构均值模型同时给每个实验处理对应的因变量建模并比较潜均值，而多指标多因模型是直接建立潜因变量对操纵变量的回归。又如，Langenberg等(2020)提出的潜重复测量方差分析(latent repeated measures analysis of variance)将单指标显性结果变量替换为多指标潜变量，提升了主效应和交互效应的检验力，放松了有关缺失和残差结构的假设，并能通过测量不变性分析验证强不变性是否满足。

工具变量是研究者不感兴趣但能够解释预测变量内生性(即预测变量与模型残差相关)，且与模型残差无关的变量。通过使用工具变量，将预测变量分解为与残差无关的外生部分及与残差相

关的内生部分，仅使用外生部分估计感兴趣的路径系数。这样可以解决内生性问题，获得对模型系数更准确的估计，从而提升 SEM 的因果推断能力(Maydeu-Olivares et al., 2020)。

尽管 SEM 有很多专用软件(如 Amos、EQS、LISREL 和 Mplus)，但传统软件也有一些局限，且很多新提出的前沿方法尚不能很快纳入传统软件中。此时，研究者自己开发的软件包就对改进传统软件局限和推动新方法应用起到重要作用。有不少方法文献致力于介绍新开发的 SEM 软件包。例如, Gonzales (2021), Rosseel (2012), Igolkina 和 Meshcheryakov (2020)分别介绍了综合性的 SEM 软件包 JMP Pro、lavaan 和 semopy, Jiang 等(2017)介绍了等效性检验软件包 equaltestMI，Zhang 等(2021)介绍了 CFA 模型修正软件包 blcfa。

### 8.3　结语

对于 SEM 的前沿方法，国内期刊文献的工作大多数属于跟踪、介绍或评论、整合，缺乏对这些方法统计学性质的深入研究。一方面，对方法之间的联系认识不够。少有研究试图厘清竞争方法之间的数理关联(如：方俊燕 等，印刷中；温涵，梁韵斯，2015)，如哪些方法是近似甚至等价的，或通过怎样的变化，不同方法可以相互转化。而此类研究在国外并不鲜见(Serang et al., 2019; Usami et al., 2015, 2019; Yuan & Deng, 2021)。此类分析有助于加深研究者对方法的全面理解和融会贯通。另一方面，方法之间比较的模拟研究也不多。比如有关拟合指数的模拟研究国外从未间断，既有模拟研究探究各种非拟合因素(如参数估计方法和信度)对拟合指数估计的影响(McNeish et al., 2018; Shi & Maydeu-Olivares, 2020)，也有研究比较新拟合指数和流行拟合指数的优劣(Counsell et al., 2020; Garnier-Villarreal & Jorgensen, 2020)。

尽管尚有不足，过去 20 年国内 SEM 方法成果仍为提升国内量化研究水平提供了强大助力，相信在方法研究和应用需求的相互促进之下，未来会有更多高质量的 SEM 方法文献出现。

## 参考文献

安敏. (2016). 结构方程模型中样本量估计的蒙特卡洛方法. *伤害医学(电子版)*, *5*(4), 45–49.

白新文, 陈毅文. (2004). 测量等价性的概念及其判定条件. *心理科学进展*, *12*(2), 231–239.

毕向阳. (2019). 基于多水平验证性因子分析的城市社区社会资本测量——实例研究及相关方法综述. *社会学研究*, *34*(6), 213–237.

卞冉, 车宏生, 阳辉. (2007). 项目组合在结构方程模型中的应用. *心理科学进展*, *15*(3), 567–576.

陈明亮. (2004). 结构方程建模方法的改进及在 CRM 实证中的应用. *科研管理*, *25*(2), 70–75.

陈楠, 刘红云. (2015). 基于增长模型的非随机缺失数据处理: 选择模型和极大似然方法. *心理科学*, *38*(2), 192–197.

陈莎, 阳庆玲, 邱佳玲, 范潇茹, 何娟, 范雄智, 郝春. (2020). 成对数据的主客体互依调节模型在公共卫生领域的应用与 SPSS 软件实现. *现代预防医学*, *47*(20), 3653–3659.

陈宇帅, 温忠麟, 顾红磊. (2015). 因子混合模型: 潜在类别分析与因子分析的整合. *心理科学进展*, *23*(3), 529–538.

程豪, 易丹辉. (2016). 偏最小二乘—二阶因子模型在综合变量构建问题的研究. *现代管理科学*, (2), 18–21.

邓居敏, 陈羽, 关颖. (2018). 基于结构方程模型的缺失数据填补方法比较. *数理医药学杂志* *31*(2), 159–161.

方杰, 邱皓政, 张敏强. (2011). 基于多层结构方程模型的情境效应分析——兼与多层线性模型比较. *心理科学进展*, *19*(2), 284–292.

方杰, 温忠麟, 欧阳劲樱, 蔡保贞. (2022). 国内调节效应的方法学研究. *心理科学进展*, *30*(8), 1703–1714.

方俊燕, 温忠麟, 黄国敏. (印刷中). 纵向关系的探究: 基于交叉滞后结构的追踪模型. *心理科学*.

方敏, 黄正峰. (2010). 结构方程模型下非正态数据的处理. *中国卫生统计*, *27*(1), 84–87.

傅珏生, 田晓明. (2004). 最大似然方法和 Bayes 方法在结构方程模型分析中的讨论(英文). *数理统计与管理*, *24*(6), 53–58.

高静. (2012). 结构方程模型中关于有序分类数据的处理方法. *统计与决策*, (18), 19–21.

高文阳, 李伟, 林伟鹏, 翁清雄, 王阳, 杨松. (2021). 潜变化分数模型在组织行为学追踪研究中的应用. *中国人力资源开发*, *38*(11), 6–25.

勾建伟, 夏业茂. (2022). 带有多重分类变量的潜变量模型的可识别性. *应用数学*, *35*(2), 302–316.

顾红磊, 温忠麟, 方杰. (2014). 双因子模型: 多维构念测量的新视角. *心理科学*, *37*(4), 973–979.

桂裕亮, 张超, 徐畅, 彭乐, 左红霞, 杨元君, 牛玉明. (2016). 结构方程模型在 Meta 分析中的应用之固定效应模型. *中国循证医学杂志*, *16*(2), 229–234.

郭庆科, 李芳, 陈雪霞, 王炜丽, 孟庆茂. (2008). 不同条件下拟合指数的表现及临界值的选择. *心理学报*, *40*(1), 109–118.

郭庆科, 王炜丽, 陈雪霞, 韩丹. (2007). 验证性因素分析中模型拟合的判断. *心理学探新*, *27*(4), 83–87.

郭晟豪, 萧鸣政. (2022). 公共管理研究中潜变量指标误用

问题及其改进策略. *华中科技大学学报(社会科学版), 36*(1),103–112.
郭小玲, 裴磊磊, 张岩波. (2009). 潜在类别模型及数据模拟分析. *数理医药学杂志, 22*(6), 631–635.
郭芸. (2005). 一般非线性结构方程模型的贝叶斯因子的计算. *苏州大学学报(自然科学版), 21*(4), 24–28.
何娟, 范雄智, 郝春. (2018). 成对数据的主客体互依模型在 MPLUS 中的实现. *现代预防医学, 45*(3), 390–400.
侯杰泰, 温忠麟, 成子娟. (2004). *结构方程模型及其应用.* 教育科学出版社.
胡传鹏, 孔祥祯, 彭凯平. (2018). 贝叶斯因子及其在 JASP 中的实现. *心理科学进展, 26*(6), 951–965.
胡鹏, 路红, 马子程. (2018). 验证性因子分析中允许误差相关的可行性与条件性. *统计与决策, 34*(19), 37–41.
黄明明. (2018). 基于混合 IRT 的潜在转变模型研究进展. *考试研究,* (1), 102–110.
黄明明. (2019). 潜在转变模型在教育测验中的应用——以英语阅读理解测验为例. *濮阳职业技术学院学报, 32*(4), 91–95.
霍映宝. (2006). LISREL 与 PLS 路径建模原理分析与比较. *统计与决策,* (20), 19–21.
贾新明. (2011). 结构方程模型评价体系的可比性问题. *数理统计与管理, 30*(2), 246–253.
贾新明, 刘亮. (2008). 结构方程模型与联立方程模型的比较. *数理统计与管理, 27*(3), 439–446.
贾跃千, 宝贡敏. (2009). 结构方程模型中的构成型测量模型研究前沿探析. *外国经济与管理, 31*(5), 52–59.
焦辛妮, 王长义, 汪东伟, 刘艳. (2015). 基于测量模型结构方程模型 GLS 与 WLS 比较. *中国公共卫生, 31*(1), 104–108.
焦辛妮, 汪东伟, 王长义, 刘艳. (2015). 结构方程模型 GLS 与 WLS 性能比较. *中国公共卫生, 31*(9), 1225–1229.
金勇进, 梁燕. (2005). 偏最小二乘(Partial Least Square)方法的拟合指标及其在满意度研究中的应用. *数理统计与管理, 24*(2), 40–44.
李冲, 刘红云. (2011). 等级数据的测量不变性检验及影响因素模拟研究. *心理科学, 34*(6), 1482–1487.
李观海, 赵丽, 邓凯升, 张瑛, 宫晓, 郜艳晖. (2020). 潜在类别与潜在类别因子分析在异质性群体分类中的比较及应用. *广东药科大学学报, 36*(1), 112–117.
李丽霞, 郜艳晖, 张敏, 张岩波. (2012). 潜变量增长曲线模型及其应用. *中国卫生统计, 29*(5), 713–716.
李丽霞, 赵丽, 周舒冬, 张敏, 郜艳晖, 张岩波. (2015). 群体异质性研究的潜变量分析方法. *中国卫生统计, 32*(4), 711–715.
李丽霞, 周舒冬, 张敏, 张岩波, 郜艳晖. (2014). 多水平模型和潜变量增长曲线模型在纵向数据分析中的应用及比较. *中华流行病学杂志, 35*(6), 741–744.
李顺勇, 岳利梅. (2017). 基于成分数据的多元结构方程模型. *河南理工大学学报(自然科学版), 36*(3), 151–156.
李晓鸿. (2012). LISREL 与 PLS 建模方法的分析与比较. *科技管理研究, 32*(20), 230–233.
李育辉, 黄飞. (2010). 成对数据分析之行动者-对象互依性模型(APIM). *心理科学进展, 18*(8), 1321–1328.
李云仙, 王学仁. (2011). 带有有序变量的结构方程模型中的模型选择. *统计与决策,* (14), 15–18.
李云仙, 王学仁. (2012). 带有缺失数据的结构方程模型中的模型选择问题. *数理统计与管理, 31*(6), 1010–1021.
李云仙, 杨爱军. (2014). 两水平结构方程模型的贝叶斯模型选择. *统计与决策,* (16), 4–9.
梁莘娅, 杨艳云. (2016). 当结构假设和分布假设不满足时的验证性因子分析: 稳健极大似然法估计和贝叶斯估计的比较研究(英文). *心理科学, 39*(5), 1256–1267.
林盛, 吴兵福, 宁禄乔, 何涛. (2006). 基于多个潜变量 PLS 算法的一种迭代方法. *系统工程学报, 21*(4), 446–448.
林小鹏, 孔丹莉, 付会斌, 丁元林. (2010). 结构方程模型的研究进展. *中国医学创新, 7*(5), 174–176.
刘畅, 伍新春. (2017). 主客体互倚性的成对模式及其检验. *心理发展与教育, 33*(1), 105–112.
柳恒超, 许燕, 王力. (2007). 结构方程模型应用中模型选择的原理和方法. *心理学探新, 27*(1), 75–78.
刘红云. (2007). 如何描述发展趋势的差异: 潜变量混合增长模型. *心理科学进展, 15*(3), 539–544.
刘红云, 李冲, 张平平, 骆方. (2012). 分类数据测量等价性检验方法及其比较: 项目阈值(难度)参数的组间差异性检验. *心理学报, 44*(8), 1124–1136.
刘红云, 骆方, 王玥, 张玉. (2012). 多维测验项目参数的估计: 基于 SEM 与 MIRT 方法的比较. *心理学报, 44*(1), 121–132.
刘金兰, 何涛, 宁禄乔, 吴兵福. (2005). 结构方程模型的偏最小二乘算法及其几何意义. *哈尔滨商业大学学报(自然科学版), 21*(6), 776–780.
刘军. (2005). 比较研究中的测量平衡性问题分析. *数理统计与管理, 24*(3), 25–31.
刘军, 吴维库. (2005). 心理测量平衡性研究与实例. *心理科学, 28*(1), 170–174.
刘小琴, 马瑞, 罗艳虹, 李治, 张春森, 张岩波. (2013). 非正态验证性因子分析在基因整体效应中的应用. *生物信息学, 11*(3), 192–195.
刘馨婷, 彭思韦, 涂冬波. (2019). 双因子模型下 CAT 测验优化设计及其效果验证. *江西师范大学学报(自然科学版), 43*(2), 128–134.
刘燕, 陈英武. (2007). 结构方程模型参数估计的 GME 方法. *国防科技大学学报, 29*(1), 116–121.
刘砚燕, 袁长蓉. (2015). 量表测量等价性及其在护理研究中的应用. *中华护理杂志, 50*(1), 110–116.
刘源. (2021). 多变量追踪研究的模型整合与拓展: 考察往复式影响与增长趋势. *心理科学进展, 29*(10), 1755–1777.
刘源, 都弘彦, 方杰, 温忠麟. (2022). 国内追踪数据分析方法研究与模型发展. *心理科学进展, 30*(8), 1734–1746.

刘源, 刘红云. (2015). 潜变量量尺的拓展及研究展望. *统计与决策*, (6), 8–12.

刘源, 骆方, 刘红云. (2014). 多阶段混合增长模型的影响因素: 距离与形态. *心理学报*, *46*(9), 1400–1412.

刘源, 赵骞, 刘红云. (2013). 多阶段增长模型的方法比较. *心理学探新*, *33*(5), 415–422.

刘玥, 刘红云. (2017). 基于双因子模型的测验总分和维度分的合成方法. *心理学报*, *49*(9), 1234–1246.

骆方, 张厚粲. (2006). 使用验证性因素分析检验测验的多维性的实验研究. *统计研究*, *23*(4), 76–79.

骆雷. (2020). PLS-SEM 多变量统计分析在赛事观众研究领域中的应用. *上海体育学院学报*, *44*(11), 86–94.

吕浥尘, 赵然. (2018). 群组发展模型——干预研究的新方法. *心理学探新*, *38*(1), 91–96.

马文超, 边玉芳, 郭雯婧, 谢敏. (2014). K-means、潜在类别模型和混合 Rasch 模型的比较. *心理学探新*, *34*(5), 431–436.

麦玉娇, 温忠麟. (2013). 探索性结构方程建模(ESEM): EFA 和 CFA 的整合. *心理科学进展*, *21*(5), 934–939.

毛秀珍, 刘欢, 唐倩. (2019). 双因子模型 MCAT 中多级评分项目选题策略的比较. *心理科学*, *42*(1), 187–193.

毛秀珍, 夏梦连, 辛涛. (2018). 全信息项目双因子分析: 模型、参数估计及其应用. *心理科学进展*, *26*(2), 358–367.

孟灿, 武俊青, 李玉艳, 周颖, 李娜, 张玉凤, 赵瑞. (2010). 潜类别分析原理及其在聚类分析中的应用. *中国卫生统计*, *27*(3), 237–239.

孟洁, 王惠文. (2009). 多元成分数据的对数衬度偏最小二乘通径分析模型. *数理统计与管理*, *28*(3), 436–442.

宁禄乔, 刘金兰. (2007). 两个潜变量的 PLS 算法模拟数据分析. *统计与决策*, (16), 159–160.

宁禄乔, 刘金兰, 吴兵福, 何涛. (2007). 结构方程模型迭代算法研究. *系统工程学报*, *22*(1), 84–87.

钱刘兰, 陈炳为, 姚宁宁, 杨建鑫, 李杰. (2015). 基于结构方程模型的 meta 分析在抑郁与社会支持关系的应用. *中国卫生统计*, *32*(1), 63–65.

秦正积, 严晓玲, 沈毅, 肖静, 何书, 任文龙.(2020). OpenBUGS 处理结构方程模型贝叶斯分析在 SAS 宏程序中的实现. *中国卫生统计*, *37*(3), 475–480.

任红梅, 王綫. (2010). 两个潜变量的模糊 PLS-结构方程模型算法求解. *统计与决策*, (7), 47–49.

任学柱, 王腾飞, Karl Schweizer. (2017). 预设路径模型及其在认知心理学研究中的应用. *心理科学进展*, *25*(10), 1675–1681.

单娜, 张笑笑. (2020). 带有测量误差协变量和不可观测反应变量下的因果推断. *吉林师范大学学报(自然科学版)*, *41*(1), 62–66.

斯介生, 肖宏伟, 蒋远营. (2014). 结构方程模型在综合评价应用中的问题和对策. *现代管理科学*, (11), 99–101.

宋琼雅, 张沥今, 潘俊豪. (2021). 贝叶斯多组比较——渐近测量不变性. *心理学探新*, *41*(1), 69–75.

宋秋月, 伍亚舟. (2017). 纵向数据潜变量增长曲线模型及其在 Mplus 中的实现. *中华流行病学杂志*, *38*(8), 1132–1135.

宋晓娟, 刘红云. (2016). 成对比较数据和排序数据的处理: 模型分析的方法. *北京师范大学学报(自然科学版)*, *52*(4), 525–531.

苏荣海, 徐茂洲. (2017). 潜在成长模型在体育运动领域的应用. *统计与信息论坛*, *32*(2), 91–95.

孙继红, 杨晓江. (2009). PLS 通径模型应用中应注意的几个问题. *统计教育*, (11), 3–10.

田飞. (2007). 用结构方程模型建构指标体系. *安徽大学学报(哲学社会科学版)*, *31*(6), 92–95.

田晓明, 傅珏生. (2005). 结构方程模型的统计方法及比较. *苏州大学学报(自然科学版)*, *21*(4), 80–85.

童乔凌, 刘天桢, 童恒庆. (2009). 结构方程模型的约束最小二乘解与确定性算法. *数值计算与计算机应用*, *30*(3), 171–180.

童乔凌, 邹雪城, 熊丽, 童恒庆. (2009). 多层与多对象结构方程模型的约束最小二乘解. *武汉理工大学学报(信息与管理工程版)*, *31*(6), 865–869.

王碧瑶, 张敏强, 张洁婷, 胡俊. (2015). 基于转变矩阵描述的个体阶段性发展: 潜在转变模型. *心理研究*, *8*(4), 36–43.

王长义, 王大鹏, 赵晓雯, 方庆伟, 刘艳. (2010). 结构方程模型中拟合指数的运用与比较. *现代预防医学*, *37*(1), 7–9.

王欢, 韩海, 蔡绍晖, 梁巧, 吴肖蒙, 王波栋, 张志敏, 柳士顺. (2012). 类别变量的结构方程模型及其应用. *中国卫生统计*, *29*(4), 522–524.

王惠文, 付凌晖. (2004). PLS 路径模型在建立综合评价指数中的应用. *系统工程理论与实践*, *24*(10), 80–85.

王惠文, 张瑛. (2007). 结构方程模型的预测建模方法. *北京航空航天大学学报*, *33*(4), 477–480.

王婧, 唐文清, 张敏强, 张文怡, 郭凯茵. (2017). 多阶段混合增长模型的方法及研究现状. *心理科学进展*, *25*(10), 1696–1704.

王凯, 陈方尧, 谭铭, 陈平雁. (2018). 一种新的评价结构方程模型拟合效果的校正拟合指数. *中国卫生统计*, *35*(3), 349–354.

王坤, 刘鹤飞, 蒋成飞. (2018). 隐马尔可夫结构方程模型及其贝叶斯估计. *数理统计与管理*, *37*(2), 272–279.

王孟成, 毕向阳. (2018). 回归混合模型: 方法进展与软件实现. *心理科学进展*, *26*(12), 2272–2280.

王孟成, 毕向阳, 叶浩生. (2014). 增长混合模型: 分析不同类别个体发展趋势. *社会学研究*, *29*(4), 220–241.

王孟成, 邓俏文. (2016). 缺失数据的结构方程建模: 全息极大似然估计时辅助变量的作用. *心理学报*, *48*(11), 1489–1498.

王孟成, 邓俏文, 毕向阳. (2017). 潜变量建模的贝叶斯方

法. *心理科学进展, 25*(10), 1682–1695.

王孟成, 邓俏文, 毕向阳, 叶浩生, 杨文登. (2017). 分类精确性指数 Entropy 在潜剖面分析中的表现: 一项蒙特卡罗模拟研究. *心理学报, 49*(11), 1473–1482.

王念新, 葛世伦, 王智宁, 吴金南. (2013). 反映式与构成式测量模型的差异: 基于 TAM 的实证研究. *系统工程理论与实践, 33*(12), 3127–3138.

王念新, 仲伟俊, 梅姝娥. (2011). 我国管理学研究中的测量模型误设及仿真分析. *管理工程学报, 25*(2), 109–117.

王若宾, 刘科成, 颜志军, 王秋明. (2014). 结构方程模型简化方法及在人机行为研究中的应用. *数学的实践与认识, 44*(3), 92–99.

王珊, 骆方, 刘红云. (2014). 迫选式人格测验的传统计分与 IRT 计分模型. *心理科学进展, 22*(3), 549–557.

王晓丽, 李西营, 邵景进. (2011). 形成性测量模型: 结构方程模型的新视角. *心理科学进展, 19*(2), 293–300.

王阳, 温忠麟, 付媛姝. (2020). 等效性检验——结构方程模型评价和测量不变性分析的新视角. *心理科学进展, 28*(11), 1961–1969.

王芝皓, 田茂再, 侯震梅. (2020). 基于偏最小二乘路径模型的分位效应测度. *系统科学与数学, 40*(4), 738–750.

韦嘉, 张春雨, 赵永萍, 张进辅. (2016). 随机截距因子分析模型在控制条目表述效应中的应用. *心理科学, 39*(4), 1005–1010.

魏修建, 郑广文. (2015). 测量不变性研究综述与理论框架. *系统工程, 33*(3), 64–71.

温聪聪, 伍伟平, 林光杰. (2018). 对齐(Alignment)——一种新的多群组分析法. *心理科学进展, 27*(1), 181–189.

温聪聪, 朱红. (2021). 随机截距潜在转变分析(RI-LTA)——个案自我转变与个案间差异的分离. *心理科学进展, 29*(10), 1773–1782.

温涵, 梁韵斯. (2015). 结构方程模型常用拟合指数检验的实质. *心理科学, 38*(4), 987–994.

温忠麟. (2017). 实证研究中的因果推理与分析. *心理科学, 40*(1), 200–208.

温忠麟, 陈虹熹, 方杰, 叶宝娟, 蔡保贞. (2022). 新世纪 20 年国内测验信度研究. *心理科学进展, 30*(8), 1682–1691.

温忠麟, 方杰, 沈嘉琦, 谭倚天, 李定欣, 马益铭. (2021). 新世纪 20 年国内心理统计方法研究回顾. *心理科学进展. 29*(8), 1331–1344.

温忠麟, 方杰, 谢晋艳, 欧阳劲樱. (2022). 国内中介效应的方法学研究. *心理科学进展, 30*(8), 1692–1702.

温忠麟, 侯杰泰. (2008). 检验的临界值: 真伪差距多大才能辨别?——评《不同条件下拟合指数的表现及临界值的选择》. *心理学报, 40*(1), 119–124.

温忠麟, 侯杰泰, 马什赫伯特. (2004). 结构方程模型检验: 拟合指数与卡方准则. *心理学报, 36*(2), 186–194.

温忠麟, 刘红云. (2020). *中介效应和调节效应: 方法及应用*. 教育科学出版社.

温忠麟, 汤丹丹, 顾红磊. (2019). 预测视角下双因子模型与高阶因子模型的一般性模拟比较. *心理学报, 51*(3), 383–391.

吴瑞林. (2010). 结构方程模型中估计收敛问题的 Monte Carlo 仿真. *统计与决策*, (6), 32–34.

吴瑞林. (2012). 基于 Tikhonov 正则化的结构方程模型参数估计方法. *统计与决策*, (20), 23–25.

吴瑞林, 涂冬波. (2013). 题目因素分析: 基于 SEM 和基于 IRT 的两类方法. *心理与行为研究, 11*(1), 124–131.

吴瑞林, 祖霁云. (2010). 多分格相关系数的估计与应用. *统计与决策*, (3), 25–28.

武淑琴, 张岩波. (2011). 结构方程模型等同性检验及其在分组比较中的应用. *中国卫生统计, 28*(3), 237–239.

武淑琴, 张岩波, 张克让, 徐勇, 孙宁. (2009). 均值结构模型在抑郁症病例-对照临床研究中的应用. *中国卫生统计*, 26(4), 352–354.

吴艳, 温忠麟. (2011). 结构方程建模中的题目打包策略. *心理科学进展, 19*(12), 1859–1867.

吴宇驹, 刘毅, 凌文辁, 路红. (2012). 二分数据的结构建模探讨——以神经质人格影响教师工作倦怠为例. *广州大学学报(自然科学版), 11*(4), 98–104.

夏业茂, 陈高燕, 刘应安. (2016). 基于多元 t-分布的隐马尔可夫潜变量模型的稳健推断. *系统科学与数学, 36*(10), 1783–1803.

肖健, 叶玲珑, 方亚. (2020). 增长混合模型在健康轨迹研究中的应用进展. *中国卫生统计, 37*(4), 637–640.

许碧云, 陈炳为, 陈启光. (2007). 结构方程模型在重复测量数据中的应用. *现代预防医学, 34*(20), 3805–3807.

许宏晨. (2010). 结构方程模型多组分析在应用语言学研究中的运用——Amos 17.0 实例演示. *中国外语教育, 3*(1), 59–67.

徐霜雪, 俞宗火, 李月梅. (2017). 预测视角下双因子模型与高阶模型的模拟比较. *心理学报, 49*(8), 1125–1136.

颜波, 褚学宁, 张磊. (2019). 结构方程模型与人工神经网络结合的用户感知建模方法. *上海交通大学学报, 53*(7), 830–837.

晏宁, 李英, 李玉磊, 郭璐, 毛志雄. (2018). 贝叶斯结构方程模型在运动与锻炼心理学中的应用. *北京体育大学学, 41*(9), 75–82.

晏宁, 毛志雄, 李英, 李玉磊, 郭璐. (2018). 小样本数据潜变量建模: 贝叶斯估计的应用. *中国体育科技, 54*(6), 52–58.

杨林山, 曹亦薇. (2012). 贝叶斯理论框架下的 2 种纵向缺失数据处理方法的比较——以潜在变量增长曲线模型为例. *江西师范大学学报(自然科学版), 36*(5), 461–465.

杨彤骥, 杨红玉, 王新海. (2010). 结构方程模型的比较: 针对用户对信息系统的接受程度. *情报杂志, 29*(3), 15–19.

叶浩生, 李明. (2014). 反映还是形成?平行还是层级?PSM 的模型建构与检验. *心理学探新, 34*(3), 265–271.

叶素静, 唐文清, 张敏强, 曹魏聪. (2014). 追踪研究中缺

失数据处理方法及应用现状分析. *心理科学进展*, *22*(12), 1985–1994.
尹奎, 彭坚, 张君. (2020). 潜在剖面分析在组织行为领域中的应用. *心理科学进展*, *28*(7), 1056–1070.
喻嘉宏, 陈小娜, 郜艳晖, 张岩波, 陈柏楠, 孔羡怡, 杨朔, 李丽霞. (2018). 潜变量增长混合模型在医学研究中的应用. *中国卫生统计*, *35*(4), 496–499.
俞立平. (2020). 学术评价中一级指标测度方法研究——结构方程降维法. *信息资源管理学报*, *10*(5), 76–84.
袁帅, 曹文蕊, 张曼玉, 吴诗雅, 魏馨怡. (2021). 通向更精确的因果分析: 交叉滞后模型的新进展. *中国人力资源开发*, *38*(2), 23–41.
曾宪华, 肖琳, 张岩波. (2013). 潜在类别分析原理及实例分析. *中国卫生统计*, *30*(6), 815–817.
翟宏堃, 李强, 魏晓薇. (2022). 结构方程模型的统计检验力分析: 原理与方法. *心理科学进展*, *30*(9), 2117−2130.
章刚勇. (2015). 结构方程模型应用: 错误设定与估计程序. *统计与信息论坛*, *30*(7), 7–15.
张建平. (1993). 一种新的统计方法和研究思路——结构方程建模述评. *心理学报*, *25*(1), 93–101.
张洁婷, 焦璨, 张敏强. (2010). 潜在类别分析技术在心理学研究中的应用. *心理科学进展*, *18*(12), 1991–1998.
张洁婷, 张敏强, 耿爽. (2012). 心理学研究中潜变量空间的特性探讨. *心理学探新*, *32*(5), 404–409.
张洁婷, 张敏强, 焦璨. (2019). 纳入式分类分析法在潜在剖面模型的后续多元回归中的应用. *心理学探新*, *39*(1), 40–46.
张洁婷, 张敏强, 焦璨, 王力田. (2013). 多水平潜在类别模型在教育评价中的应用——以英语学业能力测验为例. *教育研究与实验*, (3), 78–84.
张洁婷, 张敏强, 黎光明. (2017). 潜在剖面模型的后续分析——比较分类分析法改进后的偏差. *心理学探新*, *37*(5), 434–440.
张军. (2007). 结构方程模型构建方法比较. *统计与决策*, (18), 137–139.
张沥今, 陆嘉琦, 魏夏琰, 潘俊豪. (2019). 贝叶斯结构方程模型及其研究现状. *心理科学进展*, *27*(11), 1812–1825.
张连生, 杨洁, 向安莉. (2012). 多群组分析在医学心理学中的应用研究. *中国卫生统计*, *29*(1), 105–109.
张岩波, 刘桂芬, 卢莉, 郑建中, 徐秀娟, 药红梅. (2006). 多水平 CFA 模型在构念效度评价中的应用. *中国卫生统计*, *23*(1), 24–26.
张岩波, 刘桂芬, 徐秀娟. (2008). 多水平结构方程模型及其应用. *中国卫生统计*, *25*(2), 120–123.
张岩波, 刘桂芬, 郑建中, 徐秀娟. (2005). 验证性因子分析模型的潜变量得分及其应用. *现代预防医学*, *32*(4), 285–286.
张言彩. (2009). 结构方程模型的 Gibbs 抽样与贝叶斯估计. *统计与决策*, (6), 23–25.
张瑛, 王惠文. (2008). 结构方程模型在系统综合评估指数中的应用. *北京航空航天大学学报(社会科学版)*, *21*(1), 10–12.
赵必华. (2007). 测量等值性检验及 Amos 的实现. *中国卫生统计*, *24*(6), 659–661.
赵海峰, 万迪昉. (2003). 结构方程模型与人工神经网络模型的比较. *系统工程理论方法应用*, *12*(3), 262–269.
赵丽, 李丽霞, 周舒冬, 张岩波, 郜艳晖. (2013). 潜在剖面分析和系统聚类法比较的模拟研究. *广东药学院学报*, *29*(2), 206–209.
赵萍. (2011). 关于二阶 PLS-PM 模型中权重估计算法的探讨. *统计与决策*, (13), 4–7.
郑广文, 魏修建, 尹筱雨. (2014). 测量不变性研究的新动态. *管理现代化*, *34*(5), 123–125.
郑舒方, 张沥今, 乔欣宇, 潘俊豪. (2021). 密集追踪数据分析: 模型及其应用. *心理科学进展*, *29*(11), 1948–1969.
郑文智, 吴文毅. (2014). 结构方程模型拟合评鉴: 整体拟合、内部拟合与复核效度检验. *心理学探新*, *34*(1), 57–61.
郑显亮, 顾海根, 赵必华. (2011). 二阶因素模型的测量等价性检验——以大学生网络利他行为量表为例. *心理科学*, *34*(5), 1195–1200.
周广帅, 范冰冰, 王春霞, 游顶云, 刘言训, 薛付忠, 陈伟, 张涛. (2020). 交叉滞后路径分析在变量因果时序关系研究中的应用. *中国卫生统计*, *37*(6), 813–817.
周映雪, 欧春泉, 赵智涛, 彭成华, 陆梦洁. (2013). 基于多项相关系数的结构方程模型在大学生自我学习期望影响因素研究中的应用. *中国卫生统计*, *30*(4), 529–531.
朱利平, 刘莉. (2005). 线性结构方程参数估计的一种简单方法. *应用概率统计*, *21*(2), 161–168.
朱苗苗. (2016). 基于结构方程模型改进 ARMA 模型参数估计. *软件导刊*, *15*(9), 6–9.
庄伟卿, 刘震宇. (2013). 一种基于结构方程模型的模糊综合评价算法的改进与系统实施. *统计与决策*, (12), 11–13.
Asparouhov, T., & Muthén, B. (2021). Expanding the Bayesian structural equation, multilevel and mixture models to logit, negative-binomial, and nominal variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 622–637.
Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*(3), 507–526.
Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606.
Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*(1), 3–27.
Breitsohl, H. (2019). Beyond ANOVA: An introduction to structural equation models for experimental designs. *Organizational Research Methods*, *22*(3), 649–677.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*(2), 230–258.

Counsell, A., Cribbie, R. A., & Flora, D. B. (2020). Evaluating equivalence testing methods for measurement invariance. *Multivariate Behavioral Research, 55*(2), 312–328.

Deng, L. F., & Yuan, K.-H. (2016). Comparing latent means without mean structure models: A projection-based approach. *Psychometrika, 81*(3), 802–829.

de Roover, K., Vermunt, J. K., & Ceulemans, E. (2020). Mixture multigroup factor analysis for unraveling factor loading non-invariance across many groups. *Psychological Methods.* Advance online publication.

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood. *Psychological Methods, 25*(1), 46–70.

Gonzales, J. E. (2021). Structural equation modeling with JMP® Pro. *Measurement: Interdisciplinary Research and Perspectives, 19*(1), 80–92.

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. *European Journal of Psychological Assessment, 33*(5), 313–317.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Igolkina, A. A., & Meshcheryakov, G. (2020). semopy: A Python package for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(6), 952–963.

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(4), 555–566.

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2017). A comparison of methods for uncovering sample heterogeneity: Structural equation model trees and finite mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(2), 270–282.

Jiang, G., Mai, Y. J., & Yuan, K.-H. (2017). Advances in measurement invariance and mean comparison of latent variables: Equivalence testing and a projection-based approach. *Frontiers in Psychology, 8,* Article 1823.

Jiang, G., & Yuan, K.-H. (2017). Four new corrected statistics for SEM with small samples and nonnormally distributed data. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(4), 479–494.

Lai, K., & Zhang, X. (2017). Standardized parameters in misspecified structural equation models: Empirical performance in point estimates, standard errors, and confidence intervals. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(4), 571–584.

Langenberg, B., Helm, J. L., & Mayer, A. (2020). Repeated measures ANOVA with latent variables to analyze interindividual differences in contrasts. *Multivariate Behavioral Research.* Advance online publication.

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(1), 1–26.

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949.

Lu, Z.-H., Chow, S.-M., & Loken, E. (2016). Bayesian factor analysis as a variable-selection problem: Alternative priors and consequences. *Multivariate Behavioral Research, 51*(4), 519–539.

Mansolf, M., Jorgensen, T. D., & Enders, C. K. (2020). A multiple imputation score test for model modification in structural equation models. *Psychological Methods, 25*(4), 393–411.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320–341.

Maydeu-Olivares, A., Shi, D., & Fairchild, A. J. (2020). Estimating causal effects in linear regression models with observational data: The instrumental variables regression model. *Psychological Methods, 25*(2), 243–258.

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment, 100*(1), 43–52.

McNeish, D., & Wolf, M. G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods.* Advance online publication.

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 116–139.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of

substantive theory. *Psychological Methods, 17*(3), 313–335.

Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological Methods, 22*(4), 687–704.

Pek, J., & Wu, H. (2018). Parameter uncertainty in structural equation models: Confidence sets and fungible estimates. *Psychological Methods, 23*(4), 635–653.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Schulze, D., & Pohl, S. (2021). Finding clusters of measurement invariant items for continuous covariates. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(2), 219–228.

Serang, S., Grimm, K. J., & Zhang, Z. (2019). On the correspondence between the latent growth curve and latent change score models. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(4), 623–635.

Shi, D., & Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. *Educational and Psychological Measurement, 80*(3), 421–445.

Smid, S. C., & Winter, S. D. (2020). Dangers of the defaults: A tutorial on the impact of default priors when using Bayesian SEM with small samples. *Frontiers in Psychology, 11,* Article 3536.

Thoemmes, F., Rosseel, Y., & Textor, J. (2018). Local fit evaluation of structural equation models using graphical criteria. *Psychological Methods, 23*(1), 27–41.

Usami, S., Hayes, T., & McArdle, J. J. (2015). On the mathematical relationship between latent change score and autoregressive cross-lagged factor approaches: Cautions for inferring causal relationship between variables. *Multivariate Behavioral Research, 50*(6), 676–687.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24*(5), 637–657.

Yuan, K.-H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests. *Psychological Methods, 21*(3), 405–426.

Yuan, K.-H., & Deng, L. (2021). Equivalence of partial-least-squares SEM and the methods of factor-score regression. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 557–571.

Zhang, L., Pan, J., Dubé, L., & Ip, E. H. (2021). blcfa: An R package for Bayesian model modification in confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 649–658.

# Methodological research and model development on structural equation models in China's mainland from 2001 to 2020

WANG Yang[1], WEN Zhonglin[2], LI Wei[3,4], FANG Jie[5]

([1] *School of Public Administration, Guangdong University of Finance, Guangzhou* 510521, *China*)

([2] *School of Psychology/Center for Studies of Psychological Application, South China Normal University, Guangzhou* 510631, *China*) ([3] *School of Education Science and Technology, Northwest Minzu University, Lanzhou* 730124, *China*)

([4] *Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan* 430079, *China*)

([5] *Institute of New Development & Department of Applied Psychology, Guangdong University of Finance & Economics, Guangzhou*, 510320, *China*)

**Abstract:** In the first two decades of the 21st century, the hotspots of the methodological research on structural equation models (SEM) in China's mainland generally involve the following five aspects: model development, parameter estimation, model evaluation, measurement invariance and special data processing. Remarkably, there is more progress in model development (i.e., different variations of SEM) amongst the above aspects. After an overview of the background knowledge of these hotspots, we presented the main research topics and methodological achievements under each hotspot. We also discussed the recent progress in SEM methodological research abroad and the future research directions.

**Key words:** structural equation model; model development; parameter estimation; model evaluation; measurement invariance